Survival Analysis Using Cox Proportional Hazard Modeling for AMI(acute myocardial infarction)

Junyan Deng, MS,

Department of math and statistics, Cal State Long Beach, CA

1. Abstract

Heart Attack is a serious illness and the incidence is highly related to age. The midian survival time from KM (Kaplain Meier) estimate is approximate 2500 days. The mean survival time is 1735 days. Cox regression model and parametric distribution were used to analysis the survival rates following hospital admission for acute myocardial infarction (AMI). After checking the PH (proportional hazard) assumption, stratified Cox regression and time depending model Cox regression are to be used to adjust the model. Cpk, miord, mitype and yrgrp were not satisfied the PH assumption and were needed to be stratified on these variables. The long-term survival time was statistically related to the period of staying in the hospital. On the other hand, assuming the model fallowed a distribution and fitted the model with specific distribution. Checking the FIT statistics of AIC and SBC, the Gamma distribution was the best with smallest AIC and SBC.

2. Introduction

The aim of this study was to find out the trends over time in the incidence and analysis the survival rates following hospital admission for acute myocardial infarction (AMI). The data came from The Worcester Heart Attack Study (WHAS). Data had been collected within ten 1-year periods beginning in 1975 on all AMI patients admitted to hospitals in the Worcester, Massachusetts, Metropolitan area. The data in this paper were obtained by taking 10% random sample within 6 of the cohort years from the main data set of more than 8000 admissions, and only a small subset of variables was taken into analysis.

2. Data structure

2.1 data description

The subset of predictors used are described in Table 1 alone with their codes and values. Length of follow-up (days between hospital admission and last follow-up) was used as survival time for modelling and analysis. AGE, CPK, lenfol are continuous variables. Mitype, year and yrgrp were allocated classification variables, and the others are dummy variables.

Table 1: Description of the variables in the WHAS data set.

Data are in file WHAS.DAT
n = 481

Variable	Description	Codes / Units
ID	Identification Code	1 - 481
AGE	Age (per chart)	years
SEX	Gender	0 = Male
		1 = Female
CPK	Peak Cardiac Enzyme	International Units (iu)
SHO	Cardiogenic Shock Complications	0 = No

		1 = Yes
CHF	Left Heart Failure Complications	0 = No
		1 = Yes
MIORD	MI Order	0 = First
		1 = Recurrent
MITYPE	MI Type	1 = Q-wave
		2 = Not Q-wave
		3 = Indeterminate
YEAR	Cohort Year	1 = 1975
		2 = 1978
		3 = 1981
		4 = 1984
		5 = 1986
		6 = 1988
YRGRP	Grouped Cohort Year	1 = 1975 & 1978
		2 = 1981 & 1984
		3 = 1986 & 1988
LENSTAY	Length of Hospital Stay	Days between hospital discharge and Hospital Admission
DSTAT	Discharge Status from Hospital	0 = Alive from
		1 = Dead
LENFOL	Total Length of Follow-up	Days between Date of Last Follow-up and Hospital Hospital Admission Date
FSTAT	Status as of Last Follow-up	0 = Alive
		1 = Dead

2.2 Correlation between variables

From Table 2, mitype was not significant related to the survival time(lenfol), and that the other predictors contained information of the dependent variable(lenfol). Age sho and chf were highly correlated to each other and they were highly related to the survival time. Table 3(part of the result) showed that there were tie value in the data. Ties should be taken into consideration.

Table2

Pearson Correlation Coefficients, N = 481 Prob > |r| under H0: Rho=0								
	age	sex	cpk	sho	chf	miord	mitype	lenfol
age	1.00000	0.25260 <.0001	-0.11927 0.0088	0.15491 0.0007	0.32769 <.0001	0.07154 0.1171	0.07413 0.1044	-0.33826 <.0001
sex	0.25260 <.0001	1.00000	-0.11359 0.0127	0.05772 0.2064	0.12892 0.0046	0.01964 0.6674	0.08572 0.0603	-0.09412 0.0391
cpk	-0.11927 0.0088	-0.11359 0.0127	1.00000	0.22245 <.0001	0.01009 0.8254	-0.07988 0.0801	-0.19974 <.0001	-0.09480 0.0377
sho	0.15491 0.0007	0.05772 0.2064	0.22245 <.0001	1.00000	0.29043 <.0001	0.05352 0.2413	-0.06456 0.1574	-0.28784 <.0001
chf	0.32769 <.0001	0.12892 0.0046	0.01009 0.8254	0.29043 <.0001	1.00000	0.12789 0.0050	-0.02727 0.5508	-0.27446 <.0001
miord	0.07154 0.1171	0.01964 0.6674	-0.07988 0.0801	0.05352 0.2413	0.12789 0.0050	1.00000	0.07960 0.0812	-0.11061 0.0152
mitype	0.07413 0.1044	0.08572 0.0603	-0.19974 <.0001	-0.06456 0.1574	-0.02727 0.5508	0.07960 0.0812	1.00000	-0.04531 0.3213
lenfol	-0.33826 <.0001	-0.09412 0.0391	-0.09480 0.0377	-0.28784 <.0001	-0.27446 <.0001	-0.11061 0.0152	-0.04531 0.3213	1.00000

Table3

lenfol	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	8	1.66	8	1.66
2	16	3.33	24	4.99
3	10	2.08	34	7.07
4	7	1.46	41	8.52
5	3	0.62	44	9.15

3. Method

None parameter estimate (Kaplan- Meier) could be used to get an overall look into the survival time. The disadvantage of this method is the plot would be steps. Cox regression and its extension, stratified Cox regression and time dependent Cox regression, are semi-parameter method. Cox regression are to be used to model the survival time and some influential covariates. Before deciding the final model of Cox regression, we have to check PH (proportional hazard) assumption which we assume

that the hazard ratio is independent of survival time. If the some variables violate the PH assumption, stratified or time dependent method should be used to adjust the model. The advantage of cox regression is that we do not need to pre-assume the distribution of the data sample, specifically when we do not know or hard to know the distribution of the data. Cox regression model is semi-parametric and the plot is step in visual.

Parametric survival models are used to fit the data. We have to assume that the data follow a specific distribution and fit the data. Check the fitness of the model to decide which distribution would fit well. The advantage of this method is that the plot is smoother. But the data sometimes is hard to find out its distribution and need some guess and experience, and sometimes could get a wrong conclusion if assuming the wrong distribution and using a wrong model.

4. Survival analysis

4.1 Noneparameter method

4.1.1 KM curve and Log-rank test

The Kaplan_Meier survival function estimator is calculated as:

$$\hat{S}(t) = \prod_{t_i \leq t} \frac{n_i - d_i}{n_i},$$

Where n_i is the number of subjects at risk and d_i is the number of subjects who fail, both at time t_i.

KM (Kaplan-Meier) curve shows that the midian of survival time is approximate 2500 days. The survival rate is different in the some variables,such as age, sex, sho, chf, miord, using univeriate log-rank test. These variables are potentially used in Cox regression model.

Figure 1

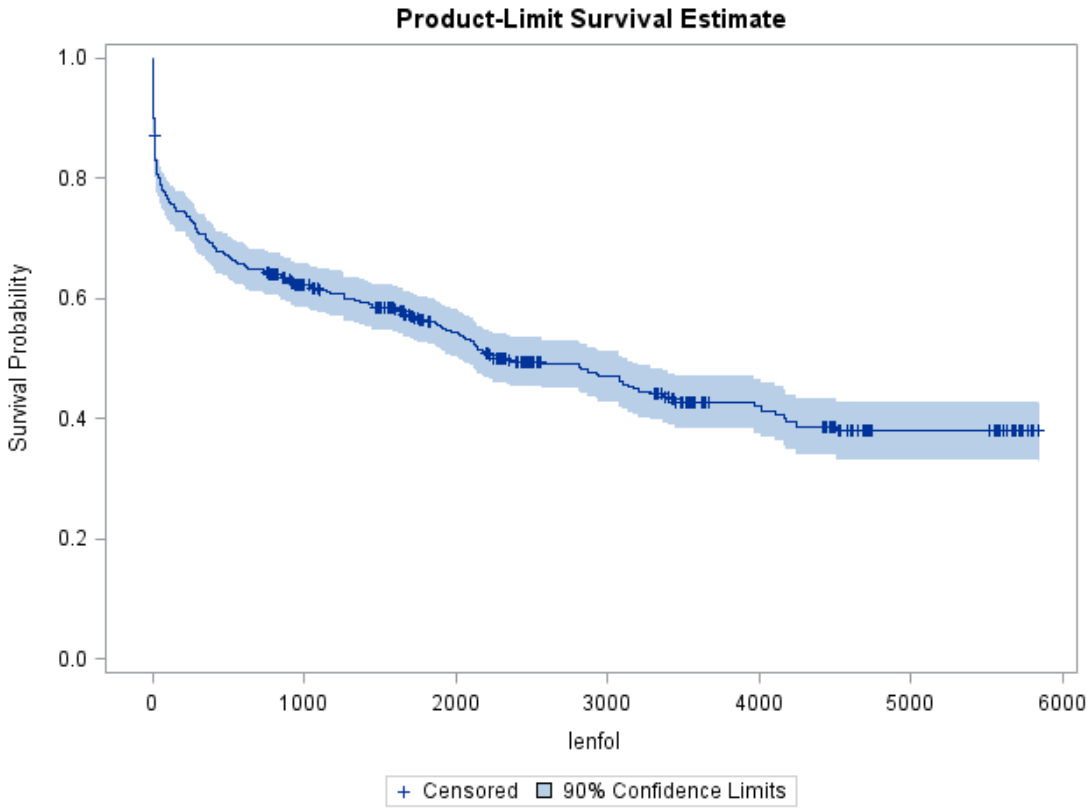


Table 3

Simple Statistics						
Variable	N	Mean	Std Dev	Sum	Minimum	Maximum
age	481	67.48441	12.68054	32460	24.00000	98.00000
sex	481	0.40333	0.49108	194.00000	0	1.00000
cpk	481	941.54262	1132	452882	10.00000	9000
sho	481	0.07900	0.27002	38.00000	0	1.00000
chf	481	0.40748	0.49188	196.00000	0	1.00000
miord	481	0.35967	0.48040	173.00000	0	1.00000
mitype	481	1.43035	0.52025	688.00000	1.00000	3.00000
lenfol	481	1735	1688	834555	1.00000	5843

Univariate Chi-Squares for the Log-Rank Test				
Variable	Test Statistic	Standard Error	Chi-Square	Pr > Chi-Square
age	-1616.0	194.9	68.7597	<.0001
sex	-23.3439	7.6220	9.3800	0.0022
cpk	-8894.7	15685.9	0.3215	0.5707
sho	-32.1880	2.1430	225.6	<.0001
chf	-61.8633	7.2955	71.9053	<.0001
miord	-25.8929	7.3844	12.2951	0.0005
mitype	5.7989	7.9661	0.5299	0.4666

4.2 Semi-parameter method

4.2.1 Cox regression

4.2.1.1 Assumming no interaction between predicters.

Assuming no interaction between predictors and using stepwise selection to get the model, including sho, age, chf, and miord. Adding sex, cpk,mitype into the model, the statiscis -2log L increase. So this is the final model of assumming no interaction between predictors. Here we keep miord in the model for checking interaction term.

Table 5 without interaction terms

Model Fit Statistics		
Criterion	Without Covariates	With Covariates
-2 LOG L	2645.732	2471.224
AIC	2645.732	2479.224
SBC	2645.732	2493.293

Analysis of Maximum Likelihood Estimates						
Parameter	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio
age	1	0.03345	0.00575	33.8735	<.0001	1.034
sho	1	1.83900	0.20593	79.7510	<.0001	6.290
chf	1	0.58309	0.14320	16.5793	<.0001	1.792
miord	1	0.25621	0.13112	3.8184	0.0507	1.292

4.2.1.2 With interaction between predictors

Stepwise selection shows the interaction of age*sho, sho*miord and sho*chf are significant. Compare the statistic, -2log L, we find that -2logL of the interacion model is less than the one without interaction. 2471.224-2451.209=20.015, the nested df=3, P-value is 0.002, less than 0.1. So we can conclude that the interaction model fits better. Here we can find that miord becomes significant when taking interaction into consideration.

Add cpk, sex and mitype into the model, the parameters do not change much, sex is not confounded with other variables. The P-value of sex is larger than 0.05 and it should be deleted in the model.

The final model is including age, sho, chf, miord, and the interction between age and sho, between sho and miord and between sho and chf.

Table 6 with interaction terms

Model Fit Statistics		
Criterion	Without Covariates	With Covariates
-2 LOG L	2645.732	2451.209
AIC	2645.732	2465.209
SBC	2645.732	2489.831

Analysis of Maximum Likelihood Estimates								
Parameter	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio	95% Hazard Ratio Confidence Limits	
age	1	0.03600	0.00610	34.8251	<.0001	1.037	1.024	1.049
sho	1	7.51344	1.56448	23.0642	<.0001	1832.508	85.381	39330.70
chf	1	0.61596	0.14581	17.8461	<.0001	1.851	1.391	2.464
miord	1	0.42601	0.13913	9.3757	0.0022	1.531	1.166	2.011
age_sho	1	-0.05096	0.01858	7.5205	0.0061	0.950	0.916	0.986
sho_chf	1	-1.22861	0.58775	4.3697	0.0366	0.293	0.092	0.926
sho_miord	1	-1.23898	0.38564	10.3221	0.0013	0.290	0.136	0.617

5. Model adequacy checking

5.1 Check PH assumption using graph

Check the PH (proportional hazard) assumption using the plot of log (-log(s (t))). If plot significant crossed, that variable violates the PH assumption, and it should be stratified later or be used time dependent variable to recode.

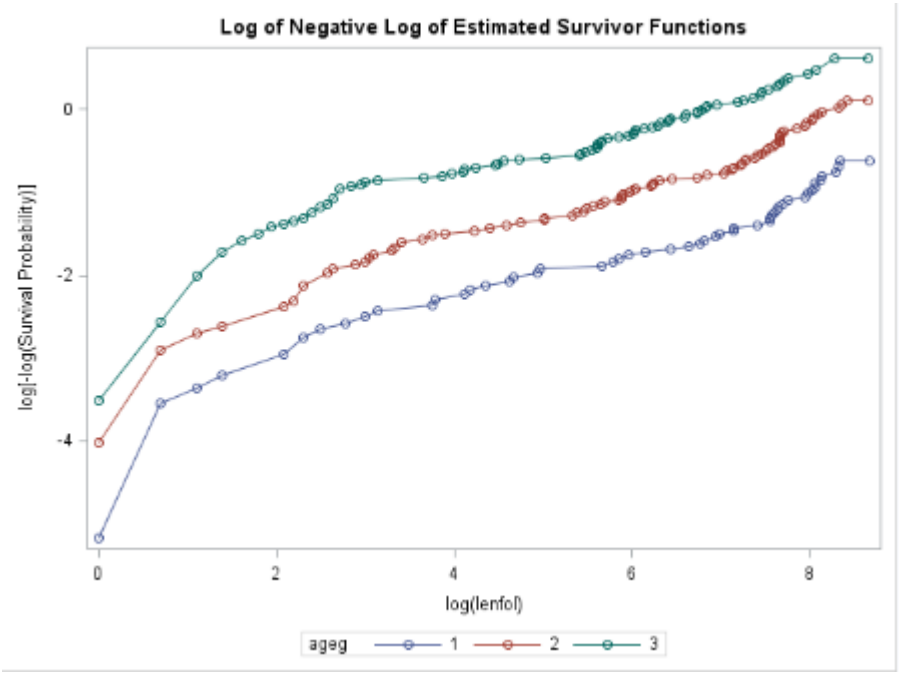
Recode age into three groups. Each group has approximate 35% percent of the total patients. Ageg=1 when age<=62. Ageg=2 when 62<age<=75. Ageg=3 when 75<age.

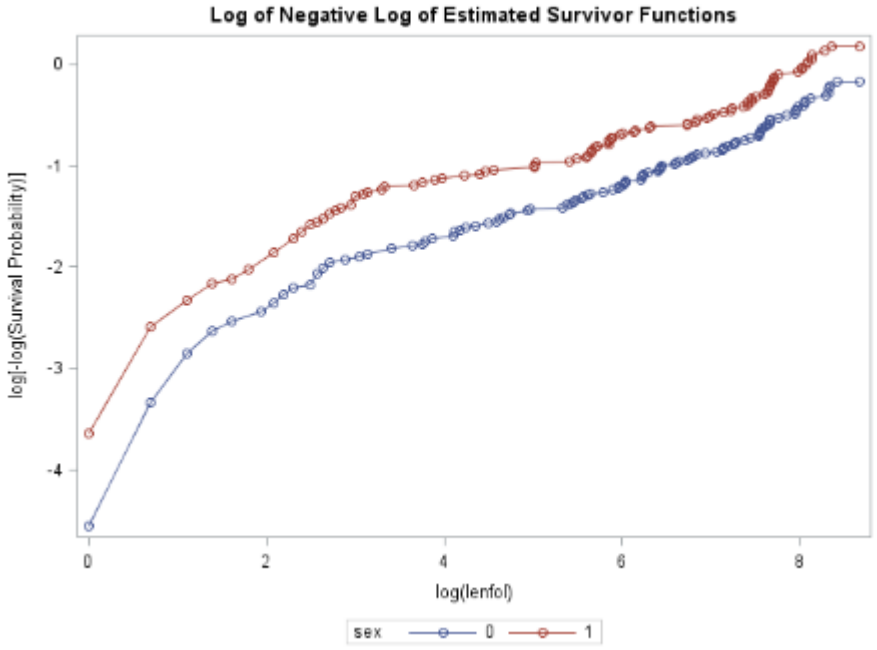
Recode cpk into three groups. Each group has approximate 35% percent of the total patients. Cpkg=1 when cpk<=98. Cpkg=2 when 98<cpk<=170. Cpkg=3 when 170<cpk.

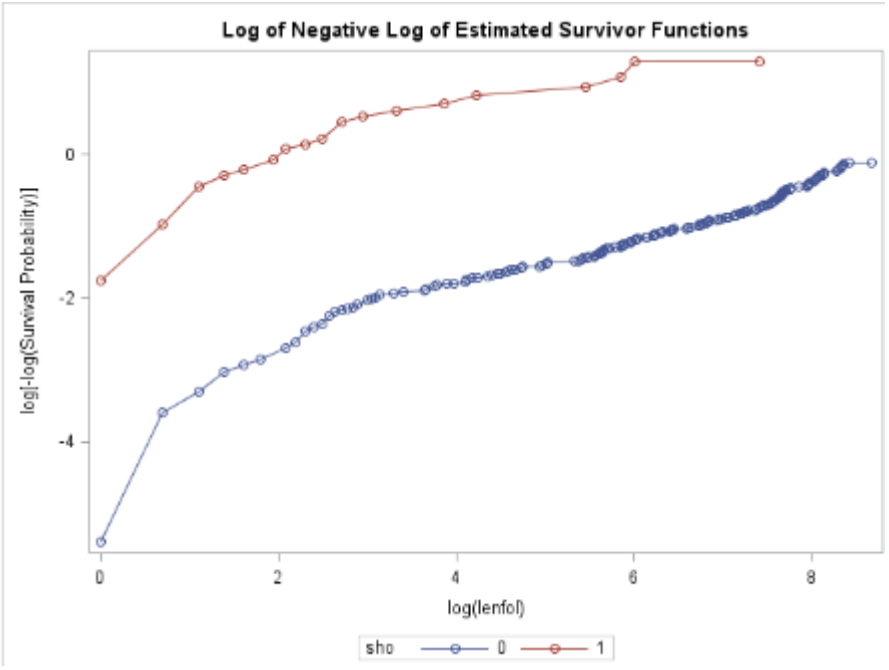
From the figures below, miord, mitype, cpkg and yrgrp violate the PH

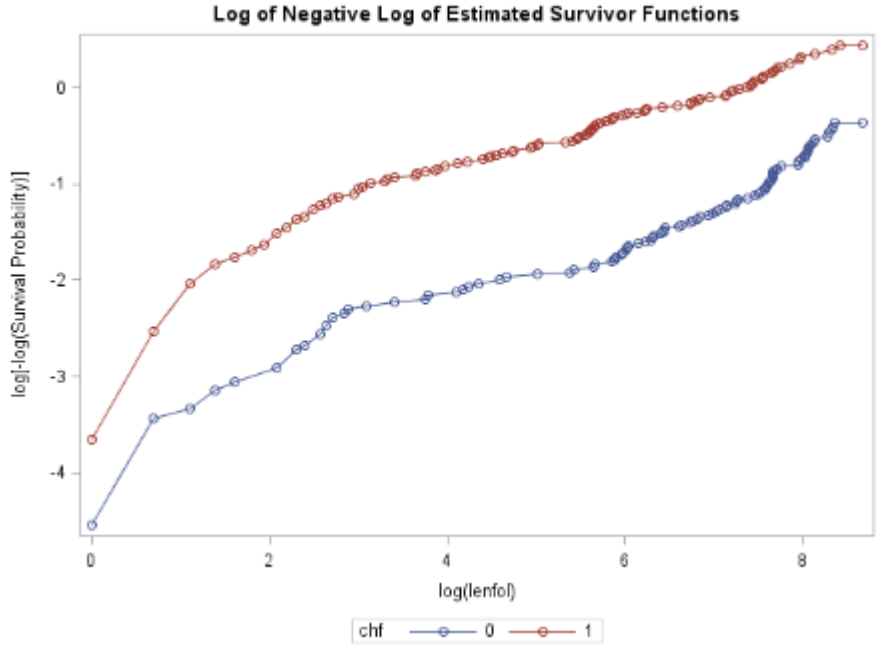
assumption, and should be stratified in Cox regression model.

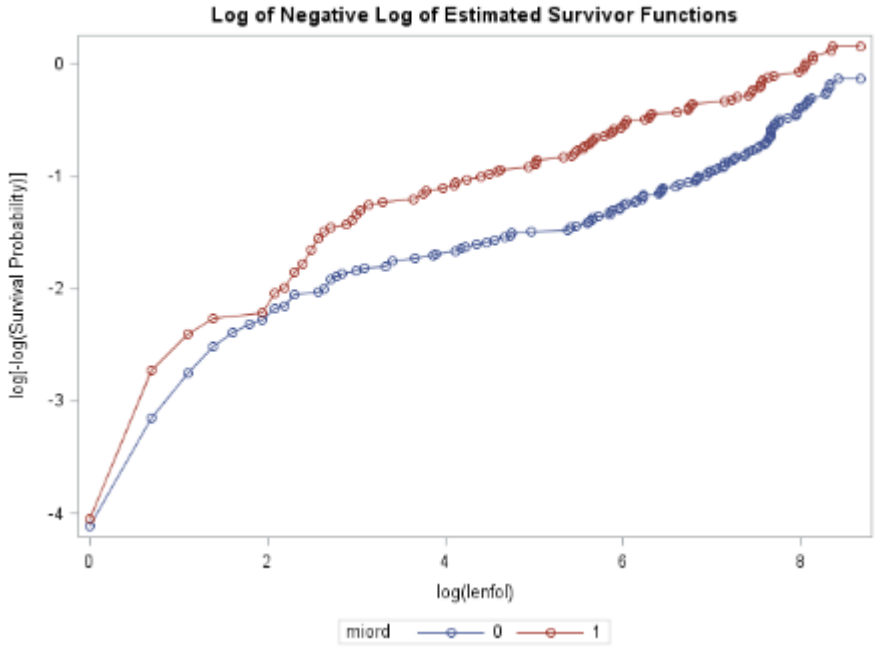
Figure 2

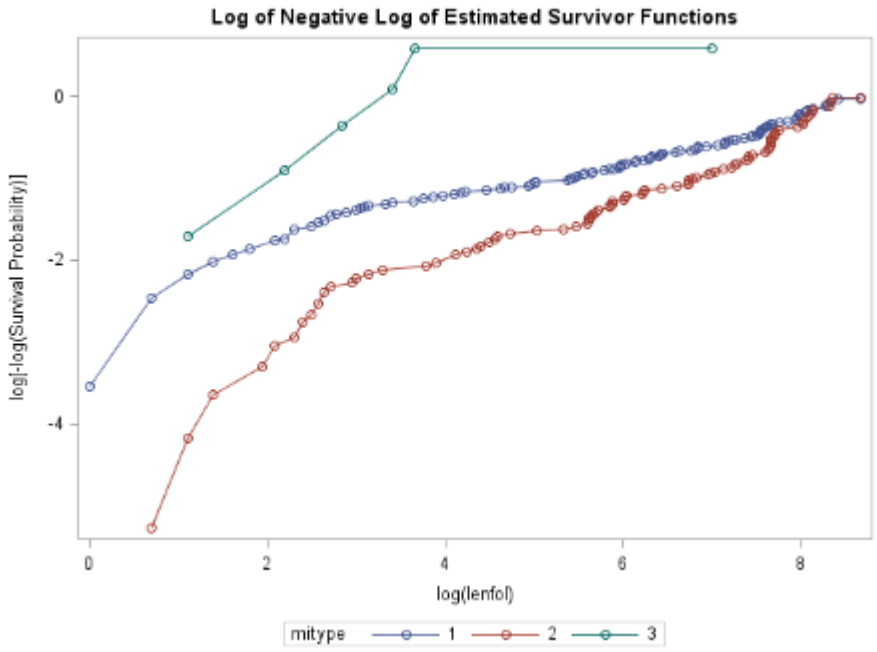

Log of Negative Log of Estimated Survivor Functions
log[-log(Survival Probability)]
0
-2
-4
0
2
4
6
8
log(lenfol)
ageg
1
2
3

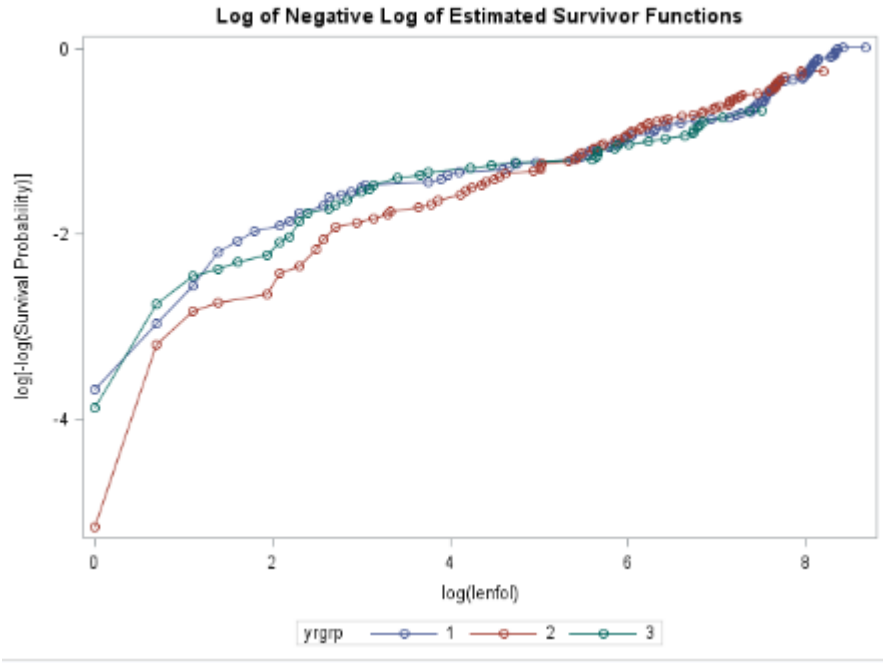

Log of Negative Log of Estimated Survivor Functions
log[-log(Survival Probability)]
0
-1
-2
-3
-4
0
2
4
6
8
log(lenfol)
sex
0
1

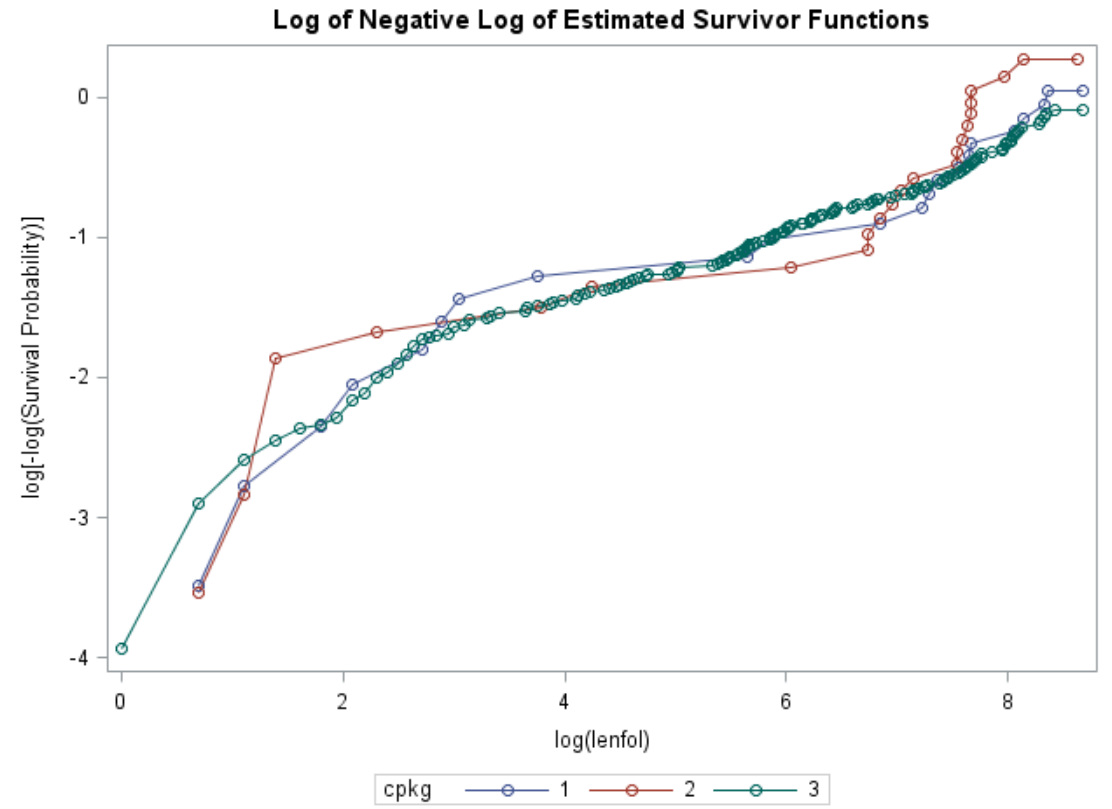

Log of Negative Log of Estimated Survivor Functions
log[-log(Survival Probability)]
0
-2
-4
0
2
4
6
8
log(lenfol)
sho
0
1


Log of Negative Log of Estimated Survivor Functions
log[-log(Survival Probability)]
0
-1
-2
-3
-4
0
2
4
6
8
log(lenfol)
chf
0
1


Log of Negative Log of Estimated Survivor Functions
log[-log(Survival Probability)]
0
-1
-2
-3
-4
0
2
4
6
8
log(lenfol)
miord
0
1


Log of Negative Log of Estimated Survivor Functions
log[-log(Survival Probability)]
0
-2
-4
0
2
4
6
8
log(lenfol)
mitype
1
2
3





5.2 Check PH assumption using assess statement

Sas provide some test to check PH assumption. Assess statement can check the functional form of continuous variable and also can check PH assumption for the final model.

```
   /*use assess statement to check age leaner function and ph
assumption.*/
 proc phreg data=whas;
     model lenfol*fstat(0)=age sho chf miord  age_sho sho_miord
sho_chf/rl ties=exact;
     age_sho=age*sho;
     sho_miord=sho*miord;
     sho_chf=sho*chf;
  assess  var=(age) ph/resample;
     run;
```

Table 7 assess statement

Supremum Test for Functional Form				
Variable	Maximum Absolute Value	Replications	Seed	Pr > MaxAbsVal
age	10.8548	1000	248585001	0.2530

Supremum Test for Proportionals Hazards Assumption				
Variable	Maximum Absolute Value	Replications	Seed	Pr > MaxAbsVal
age	0.6430	1000	248585001	0.8050
sho	3.5206	1000	248585001	0.8640
chf	1.3698	1000	248585001	0.0580
miord	1.5418	1000	248585001	0.0200
age_sho	2.5760	1000	248585001	0.9520
sho_miord	1.3560	1000	248585001	0.2050
sho_chf	1.5698	1000	248585001	0.7020

From Table 7, we can find that P-value is greater than 0.1, we can conclude the linear function of age is reasonable. The table also show that miord violate PH assumption, under the significant leva alpha=0.05.

6. Stratified Cox regression

6.1 Stratified on cpkg, miord and mitype assuming they do not interact with other variables.

The interaction between age and sho becomes insignificant and the coefficient of sho and the coefficient of the interaction between sho and chf change a lot. The statistics AIC and SBC decrease much. We conclude that stratified on miord, mitype and improves the model.

Table 8

stratified on miord, mitype and assuming they are not interact with covariates

Model Fit Statistics		
Criterion	Without Covariates	With Covariates
-2 LOG L	1773.411	1617.389
AIC	1773.411	1627.389
SBC	1773.411	1644.977

Analysis of Maximum Likelihood Estimates								
Parameter	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio	95% Hazard Ratio Confidence Limits	
age	1	0.03447	0.00621	30.8223	<.0001	1.035	1.023	1.048
sho	1	5.67816	1.71226	10.9970	0.0009	292.410	10.198	8384.352
chf	1	0.61047	0.15267	15.9882	<.0001	1.841	1.365	2.484
age_sho	1	-0.03209	0.02044	2.4634	0.1165	0.968	0.930	1.008
sho_chf	1	-1.58345	0.60520	6.8456	0.0089	0.205	0.063	0.672

Delete age_sho

Model Fit Statistics		
Criterion	Without Covariates	With Covariates
-2 LOG L	1773.411	1619.820
AIC	1773.411	1627.820
SBC	1773.411	1641.890

Analysis of Maximum Likelihood Estimates								
Parameter	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio	95% Hazard Ratio Confidence Limits	
age	1	0.03157	0.00591	28.5714	<.0001	1.032	1.020	1.044
sho	1	3.11411	0.56218	30.6842	<.0001	22.513	7.480	67.759
chf	1	0.63212	0.15200	17.2936	<.0001	1.882	1.397	2.535
sho_chf	1	-1.40777	0.59100	5.6740	0.0172	0.245	0.077	0.779

Table 9

stratified on miord, mitype, cpkg, and yrgrp assuming no interaction between stratified variables and covariates

Model Fit Statistics		
Criterion	Without Covariates	With Covariates
-2 LOG L	1379.098	1230.048
AIC	1379.098	1240.048
SBC	1379.098	1257.635

Analysis of Maximum Likelihood Estimates								
Parameter	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio	95% Hazard Ratio Confidence Limits	
age	1	0.03638	0.00648	31.5064	<.0001	1.037	1.024	1.050
sho	1	5.43106	1.76852	9.4308	0.0021	228.392	7.134	7312.253
chf	1	0.59324	0.15490	14.6676	0.0001	1.810	1.336	2.452
age_sho	1	-0.03225	0.02098	2.3622	0.1243	0.968	0.929	1.009
sho_chf	1	-1.20948	0.69308	3.0453	0.0810	0.298	0.077	1.161

Delete age_sho and sho_chf

Model Fit Statistics		
Criterion	Without Covariates	With Covariates
-2 LOG L	1379.098	1234.522
AIC	1379.098	1240.522
SBC	1379.098	1251.074

Analysis of Maximum Likelihood Estimates								
Parameter	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio	95% Hazard Ratio Confidence Limits	
age	1	0.03423	0.00616	30.8702	<.0001	1.035	1.022	1.047
sho	1	1.91197	0.23662	65.2922	<.0001	6.766	4.255	10.759
chf	1	0.56695	0.15149	14.0053	0.0002	1.763	1.310	2.372

6.1.1 Stratified on miord, mitype, cpkg, and yrgrp assuming no interaction between stratified variables and covariates

Adding yrgrp into the stratified groups. The interaction between age and sho, and the interaction between sho and chf becomes insignificant. And the statistics AIC and SBC decrease much, so we can conclude that adding yrgrp into the stratified group improve the model.

6.2 stratified on miord, mitype,cpkg, and yrgrp assuming there are interaction between stratified variables and covariates

Put the interaction terms between stratified variables and covariates into the model to check whether the interaction significant or not. Table 10 shows that the interaction between yrgrp and sho is significant and the interaction between sho and miord is significant too. The main effect of sho is strongly affected.

Table 10

Analysis of Maximum Likelihood Estimates								
Parameter	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio	95% Hazard Ratio Confidence Limits	
age	1	0.01343	0.04443	0.0913	0.7625	1.014	0.929	1.106
sho	1	17.28756	708.61661	0.0006	0.9805	32202803	0.000	.
chf	1	-1.22155	0.97866	1.5580	0.2120	0.295	0.043	2.007
age_miord	1	-0.00462	0.01330	0.1208	0.7281	0.995	0.970	1.022
age_yrgrp	1	-0.0006320	0.00908	0.0048	0.9445	0.999	0.982	1.017
age_mitype	1	0.02044	0.01309	2.4390	0.1184	1.021	0.995	1.047
age_cpkg	1	-0.00229	0.01460	0.0246	0.8754	0.998	0.970	1.027
sho_yrgrp	1	-0.89949	0.40112	5.0285	0.0249	0.407	0.185	0.893
sho_mitype	1	0.62886	0.62067	1.0265	0.3110	1.875	0.556	6.330
sho_miord	1	-2.10545	0.54630	14.8534	0.0001	0.122	0.042	0.355
sho_cpkg	1	-4.30475	236.20538	0.0003	0.9855	0.014	0.000	1.54E199
chf_miord	1	0.40860	0.31835	1.6474	0.1993	1.505	0.806	2.808
chf_yrgrp	1	0.15381	0.21477	0.5129	0.4739	1.166	0.766	1.777
chf_mitype	1	0.18339	0.31404	0.3410	0.5592	1.201	0.649	2.223
chf_cpkg	1	0.39857	0.31570	1.5939	0.2068	1.490	0.802	2.766

Delete insignificant terms and get the table we get that the interaction between sho and yrgrp becomes insignificant. This maybe cause by the main effect of sho. Comparing the statistics AIC and SBC with the ones of the model of assuming no interaction between stratified variables and covariates, they decrease much. And the coefficient of sho change a lot. So we conclude that the interaction between sho and miord improve the model, and conclude that the coefficient of the parameter are constant across cohort yrgrp when stratified on yrgrp, cpkg, miord and mitype.

Table 11 delete insignificant interaction terms from table 10.

Model Fit Statistics		
Criterion	Without Covariates	With Covariates
-2 LOG L	1379.098	1217.979
AIC	1379.098	1227.979
SBC	1379.098	1245.566

Analysis of Maximum Likelihood Estimates								
Parameter	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio	95% Hazard Ratio Confidence Limits	
age	1	0.03185	0.00616	26.7418	<.0001	1.032	1.020	1.045
sho	1	4.39998	0.94457	21.6987	<.0001	81.449	12.790	518.687
chf	1	0.59409	0.15139	15.3987	<.0001	1.811	1.346	2.437
sho_yrgrp	1	-0.63395	0.35736	3.1469	0.0761	0.530	0.263	1.069
sho_miord	1	-1.72899	0.49415	12.2426	0.0005	0.177	0.067	0.467

Delete sho_yrgrp

Model Fit Statistics		
Criterion	Without Covariates	With Covariates
-2 LOG L	1379.098	1221.312
AIC	1379.098	1229.312
SBC	1379.098	1243.382

Analysis of Maximum Likelihood Estimates								
Parameter	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio	95% Hazard Ratio Confidence Limits	
age	1	0.03256	0.00617	27.8993	<.0001	1.033	1.021	1.046
sho	1	2.90107	0.36483	63.2335	<.0001	18.194	8.900	37.193
chf	1	0.59659	0.15060	15.6932	<.0001	1.816	1.352	2.439
sho_miord	1	-1.70774	0.47937	12.6913	0.0004	0.181	0.071	0.464

6.4 Time dependent Cox regression model

It is quite possible that the long term survival in the cohort group, yrgrp, is due to the different length of staying in the hospital. Take time dependent variable x into the model to check this suspect. Define x (t)=0 when lenfol is less than or equal to lenstay, x(t)=1 when lenfol is greater than lenstay.

Table 12 show that x is significant. We should take x time dependent variable into consideration.

Table 12 time dependent variable into the model

Model Fit Statistics		
Criterion	Without Covariates	With Covariates
-2 LOG L	1379.098	1212.682
AIC	1379.098	1222.682
SBC	1379.098	1240.269

Analysis of Maximum Likelihood Estimates								
Parameter	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio	95% Hazard Ratio Confidence Limits	
age	1	0.03231	0.00618	27.2983	<.0001	1.033	1.020	1.045
sho	1	2.73648	0.36796	55.3067	<.0001	15.433	7.503	31.743
chf	1	0.56564	0.15169	13.9049	0.0002	1.761	1.308	2.370
sho_miord	1	-1.59416	0.48068	10.9990	0.0009	0.203	0.079	0.521
x	1	-1.21339	0.43531	7.7697	0.0053	0.297	0.127	0.698

From now, the final Cox regression model is stratified on miord, mitype, ckpg and yrgrp, including age, sho, chf, x (time dependent variable) and the interaction between sho and miord. We can conclude that the long term survival due to the long time staying in the hospital.

The HR (age change in 1 year) is 1.033, 95% confidence interval ranging from 1.020 to 1.045. The patients one 1 year older are 3.3% more likely at risk to die.

When miord is 0, the HR (sho 1 to sho 0) is 15.443, 95% confidence interval ranging from 7.503 to 31.743. The patients with sho 1 are 15.433 times than the patient with sho 0 more likely at risk to die.

When miord is 1, the HR (sho 1 to sho 0) is 1.14.

The HR (chf 1 to chf 0) is 1.641, 95% confidence interval ranging from 1.308 to 2.370. The patients with chf 1 are 64.1% more than the patients with chf 0 to die.

For the time dependent variable x, the HR(x=1 x=0) is 0.297, 95% confidence interval ranging from 0.127 to 0.698. The patients staying in the hospital with long time survive 3.37 (1/0.297) times longer than the staying short one.

7.1 parameter

Assuming the survival time satisfied specified distribution of exponential, Weibull, lognormal, log-logistic, and gamma, fit the model and find out that Weibull, log-logistic, lognormal, and gamma fit good, comparing AIC and SBC. These model all show that sex, cpk, and mitype are insignificant. Delete these there insignificant variables and fit again. The AIC and SBC from these models are very closed, but Gamma fits best with the smallest AIC and SBC.

Table 13 compare fit statistics

Weibull

Fit Statistics	
-2 Log Likelihood	1553.250
AIC (smaller is better)	1565.250
AICC (smaller is better)	1565.427
BIC (smaller is better)	1590.305

Log-logistic

Fit Statistics	
-2 Log Likelihood	1547.050
AIC (smaller is better)	1559.050
AICC (smaller is better)	1559.228
BIC (smaller is better)	1584.106

Log nomal

Fit Statistics	
-2 Log Likelihood	1546.876
AIC (smaller is better)	1558.876
AICC (smaller is better)	1559.053
BIC (smaller is better)	1583.931

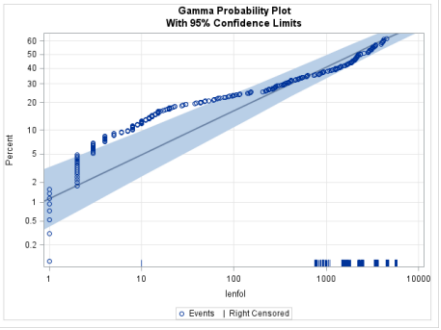
Gamma

Fit Statistics	
-2 Log Likelihood	1543.671
AIC (smaller is better)	1557.671
AICC (smaller is better)	1557.908
BIC (smaller is better)	1586.902

Table 13 parameter estimates of gamma model.

Gamma

Analysis of Maximum Likelihood Parameter Estimates							
Parameter	DF	Estimate	Standard Error	95% Confidence Limits		Chi-Square	Pr > ChiSq
Intercept	1	14.2969	0.9499	12.4351	16.1587	226.52	<.0001
age	1	-0.0758	0.0134	-0.1021	-0.0494	31.81	<.0001
sho	1	-4.6040	0.5119	-5.6073	-3.6007	80.89	<.0001
chf	1	-1.4582	0.3435	-2.1315	-0.7850	18.02	<.0001
miord	1	-0.8156	0.3229	-1.4483	-0.1828	6.38	0.0115
Scale	1	2.7924	0.2281	2.3793	3.2774		
Shape	1	0.3796	0.2032	-0.0186	0.7779		



All the parameters are negative, mean that they are all negative associate with the survival time.

8 conclussion

Heart attack is a serious illness and the midian survival time of paitients is approxmiate 2500 days. Age, sho, chf and mirod are highly related with the survival time. The survival time is

statistically the same in gender. PH assumption was not satified on cpk, miord,mitype and yrgrp. Stratified on these variables and find out that miord interacts with sho. It need further analysis on the interaction between miord and sho. The long term survival time is related to the time staying in the hospital. This suggests patients should addmit as earlly as possible and stay as longer as possible. Age, sho (Cardiogenic Shock Complications), chf (Left Heart Failure Complications), miord (MI Order) are negative associated with the survival time based on the analysis from parameter regression model. As age increases the shorter the suvival time is. Cardiogenic shock complications decreases the survival time. Patiets who have left heart failure complications live shorter survival time. Recurrence of MI would decrease the survival time.

Appendix

Sas code

```
data whas;
  infile "F:\course\590\hw\hw2\whas.dat";
  input id age sex cpk sho chf miord mitype year yrgrp lenstay dstat
lenfol fstat;
run;

proc lifetest data=whas plot=(d ls h);
time age;
test miord;
strata miord;
run;

proc corr data=whas plots(maxpoints=none)=matrix;
var age sex cpk sho chf miord mitype lenfol;
run;

/*KM estimate*/
proc lifetest data=whas plots=survival;
time lenfol*fstat(0);
test age sex cpk sho chf miord mitype year;
run;
/*check ph assumption on sex*/
proc lifetest data=whas conftype=linear plots=(survival(cl) lls ls)
```

```
outsurv=a;
time lenfol*fstat(0);
strata sex;
run;


/*get how many ties to see whether should take ties into
consideration*/
proc freq data=whas;
table lenfol;
run;

/*stepwise selection without interaction*/
proc phreg data=whas;
    model lenfol*fstat(0)=age sex cpk sho chf miord
mitype/selection=stepwise ties=efron sle=0.25 sls=0.1 ties=exact;

run;
proc phreg data=whas;
   model lenfol*fstat(0)=age sho chf miord/ties=exact;
run;
/*from stepwise we can see that cpk is not contribute much, add sex
and mitype into the model to see if something confounding.*/
proc phreg data=whas;
   model lenfol*fstat(0)=age sho chf miord sex/ties=exact;
    run;
proc phreg data=whas;
   model lenfol*fstat(0)=age sho chf miord mitype/ties=exact;
    run;


/*stepwise selection with interaction*/
proc phreg data=whas;
    model lenfol*fstat(0)=age | sho |chf| miord /selection=stepwise
ties=efron sle=0.5 sls=0.05 ties=exact;
run;
proc phreg data=whas;
    model lenfol*fstat(0)=age sho chf miord  age_sho sho_chf
sho_miord /rl ties=exact;
    age_sho=age*sho;
    sho_chf=sho*chf;
    sho_miord=sho*miord;
    run;
```

```
/*check ph assumption by plot*/
proc lifetest data=whas conftype=linear plots=(lls s ls) ;
time lenfol*fstat(0);
strata sho;
run;
proc lifetest data=whas conftype=linear plots=(lls s ls) ;
time lenfol*fstat(0);
strata chf;
run;
proc lifetest data=whas conftype=linear plots=(lls s ls) ;
time lenfol*fstat(0);
strata miord;
run;
proc freq data=whas;
table age;
run;
proc freq data=whas;
table cpk;
run;
data whas2;
set whas;
if age<=62 then ageg=1;
   else if 62<age<=75  then ageg=2 ;
else if 75<age then ageg=3 ;
run;
data whas2;
set whas2;
if cpk<=98 then cpkg=1;
    else if 98<cpk<=170 then cpkg=2;
        else if 170<cpk then cpkg=3;
run;
proc lifetest data=whas2 conftype=linear plots=(lls s ls);
time lenfol*fstat(0);
strata ageg ;
run;
proc lifetest data=whas2 plots=(lls s ls);
time lenfol*fstat(0);
strata cpkg;
run;
proc lifetest data=whas plots=(lls s ls);
time lenfol*fstat(0);
strata year;
run;
proc lifetest data=whas plots=(lls s ls);
```

```
time lenfol*fstat(0);
strata yrgrp;
run;


/*use assess statement to check age leaner function and ph
assumption.*/
 proc phreg data=whas;
     model lenfol*fstat(0)=age sho chf miord  age_sho sho_miord
sho_chf/rl ties=exact;
     age_sho=age*sho;
     sho_miord=sho*miord;
     sho_chf=sho*chf;
  assess  var=(age ) ph/resample;
     run;


/*assuming miord is not interact with other variables.*/
 proc phreg data=whas2;
     model lenfol*fstat(0)=age sho chf  age_sho sho_chf/rl ties=exact;
     age_sho=age*sho;
     sho_chf=sho*chf;
  strata miord mitype cpkg;
     run;
proc phreg data=whas2;
     model lenfol*fstat(0)=age sho chf  age_sho sho_chf/rl ties=exact;
     age_sho=age*sho;
     sho_chf=sho*chf;
  strata miord mitype cpkg yrgrp;
     run;
proc phreg data=whas2;
   model lenfol*fstat(0)=age sho chf/rl ties=exact;
   strata miord mitype cpkg yrgrp;
   run;
/*check interaction with other viriables*/
 proc phreg data=whas2;
     model lenfol*fstat(0)=age sho chf age_miord age_yrgrp age_mitype
age_cpkg
           sho_yrgrp sho_mitype sho_miord sho_cpkg
           chf_miord chf_yrgrp chf_mitype chf_cpkg/rl ties=exact;
     age_miord=age*miord;
     age_yrgrp=age*yrgrp;
     age_mitype=age*mitype;
     age_cpkg=age*cpkg;
```

```
    sho_miord=sho*miord;
    sho_yrgrp=sho*yrgrp;
    sho_mitype=sho*mitype;
    sho_cpkg=sho*cpkg;
    chf_miord=chf*miord;
    chf_yrgrp=chf*yrgrp;
    chf_mitype=chf*mitype;
    chf_cpkg=chf*cpkg;
  strata miord yrgrp mitype cpkg;
    run;


/*check interaction with other viriables*/
 proc phreg data=whas2;
    model lenfol*fstat(0)=age sho chf
          sho_yrgrp  sho_miord /rl ties=exact;
    sho_miord=sho*miord;
    sho_yrgrp=sho*yrgrp;
  strata miord yrgrp mitype cpkg;
    run;
 proc phreg data=whas2;
    model lenfol*fstat(0)=age sho chf
           sho_miord /rl ties=exact;
    sho_miord=sho*miord;
  strata miord yrgrp mitype cpkg;
    run;

proc phreg data=whas2;
    model lenfol*fstat(0)=age sho chf
           sho_miord x/rl ties=exact;
    sho_miord=sho*miord;
if lenstay<lenfol then x=0;
   else x=1;
strata yrgrp miord mitype cpkg;
run;

proc lifereg data=whas;
    model lenfol*fstat(0)=age sex cpk sho chf miord
mitype/dist=exponential;
probplot;
run;
proc lifereg data=whas;
    model lenfol*fstat(0)=age sex cpk sho chf miord
mitype/dist=weibull;
```

```
probplot;
run;
proc lifereg data=whas;
    model lenfol*fstat(0)=age sex cpk sho chf miord
mitype/dist=llogistic;
probplot;
run;
proc lifereg data=whas;
    model lenfol*fstat(0)=age sex cpk sho chf miord
mitype/dist=logistic;
probplot;
run;
proc lifereg data=whas;
    model lenfol*fstat(0)=age sex cpk sho chf miord
mitype/dist=lnormal;
probplot;
run;

proc lifereg data=whas;
    model lenfol*fstat(0)=age sex cpk sho chf miord
mitype/dist=normal;
probplot;
run;
proc lifereg data=whas;
    model lenfol*fstat(0)=age sex cpk sho chf miord
mitype/dist=gamma;
probplot;
run;
```